

Products of Stochastic Matrices with Aperiodic Core

Thomas Nowak

Laboratoire d’Informatique, École polytechnique, 91128 Palaiseau, France

Abstract

We show a new convergence theorem for infinite products of stochastic matrices. It generalizes the popular condition of a positive diagonal of all factors to the existence of what we call an aperiodic core. An aperiodic core is a common aperiodic digraph without sinks that is a subgraph of the digraph described by the factor matrices. Our result generalizes both a theorem of Wolfowitz and convergence theorems in the application domain of asymptotic consensus.

1 Introduction

A stochastic matrix is a real matrix with nonnegative entries whose row sums are equal to 1. Long and infinite products of stochastic matrices have found numerous applications, in particular in computer science for multi-agent systems whose agents start with a private value and repeatedly form averages of perceived values of others. These types of multi-agent systems are not only used in computer networks, but have also been found to model various physical and biological phenomena like the behavior of bird flocks [1, 3].

In these applications, it is of interest to know under which conditions an infinite product of stochastic matrices converges. Also of interest are conditions under which the limiting matrix has rank 1, for this ensures asymptotic agreement amongst the agents.

The first convergence result for products of stochastic matrices is the Perron-Frobenius theorem, which states that the powers of an ergodic stochastic matrix converge to a rank 1 stochastic matrix. Ergodicity here is equivalent to the fact that the matrix has a power that is positive, i.e., all of its entries are positive. This result was first generalized to a non-constant product of matrices by Wolfowitz [6]:

Theorem 1 (Wolfowitz, 1963). *Let \mathcal{M} be a finite set of stochastic $n \times n$ matrices with the property that every finite product of matrices in \mathcal{M} is ergodic, and let $A(k)$ be a sequence of elements of \mathcal{M} . Then the sequence of products $A(k) \cdot A(k-1) \cdots A(1)$ converges to a rank 1 stochastic matrix.*

The strict finiteness and ergodicity conditions in Wolfowitz’ theorem were found to be inappropriate for the mentioned applications. Subsequently, Theorem 1 was extended in several directions (see, for example, [1, Theorem 2] or [3, Section II.G]). However, no direct generalization of Wolfowitz’ theorem was obtained. These results required a constant lower bound on positive matrix entries and an eventual positivity assumption in the product. These conditions are also fulfilled in Wolfowitz theorem. Additionally, they required a positive diagonal and pseudo-symmetry of the matrices. (More specifically, a pseudo-symmetric matrix A is one such that $A_{i,j}$ is positive if and only if $A_{j,i}$ is.)

These results were recently generalized by Touri and Nedić [4] as part of a result on random products of stochastic matrices. They replaced pseudo-symmetry by the more general property of complete reducibility. A matrix is *completely reducible* if it is not permutation similar to a matrix of the form

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where both A and C are quadratic and $B \neq 0$. A non-probabilistic version of their theorem is the following:

Theorem 2 (Touri and Nedić, 2011). *Let $A(k)$ be a sequence of $n \times n$ stochastic matrices with the following properties:*

1. *Positive entries are lower bounded, i.e., there exists an $\alpha > 0$ such that $A_{i,j}(k) > 0$ implies $A_{i,j}(k) \geq \alpha$ for all i, j , and k .*

2. *Eventual positivity:*

For every k there exists a $K \geq k$ such that the matrix $\sum_{k'=k}^K A(k') \cdot A(k'-1) \cdots A(k)$ is positive.

3. *Every $A(k)$ is completely reducible.*

4. *All diagonal entries $A_{i,i}(k)$ are positive.*

Then the sequence of products $A(k) \cdot A(k-1) \cdots A(1)$ converges to a stochastic matrix of rank 1.

If the eventual positivity condition (2) is omitted, convergence still holds, but the limit is not guaranteed to have rank 1. There also exists an analogous variant of our main theorem when this condition is dropped.

The purpose of this paper is to generalize Theorem 2 by replacing condition (4) of positive diagonal entries by the more general requirement for the existence of an *aperiodic core*. The notion of an aperiodic core is a graph-theoretical one which we detail now:

To every stochastic $n \times n$ matrix A , we assign the digraph $G(A)$ with node set $[n] = \{1, 2, \dots, n\}$ containing an edge (i, j) if and only if the matrix entry $A_{i,j}$ is positive. A matrix is completely reducible if and only if its digraph does not contain edges between its strongly connected components. A strongly connected digraph is *aperiodic* if the lengths of the cycles it contains are coprime. A not strongly connected digraph is aperiodic if all its strongly connected components are. Clearly, a digraph containing self-loops at all nodes is aperiodic. This corresponds to the case of a positive diagonal. We will use Wielandt's theorem [5] on the existence of walks with prescribed lengths in aperiodic digraphs to replace the self-loops, which have an exponent equal to zero, by an arbitrary aperiodic core.

The following theorem is the main result of this paper. It generalizes at the same time Theorem 1 and Theorem 2. A *sink* is a node without outgoing edges.

Theorem 3. *Let $A(k)$ be a sequence of $n \times n$ stochastic matrices with the following properties:*

1. *Positive entries are lower bounded.*

2. *Eventual positivity:*

For every k there exists a $K \geq k$ such that the matrix $\sum_{k'=k}^K A(k') \cdot A(k'-1) \cdots A(k)$ is positive.

3. *Every $A(k)$ is completely reducible.*

4. *The sequence has an aperiodic core, i.e., there exists an aperiodic digraph H with node set $[n]$ and without sinks such that H is a subgraph of every $G(A(k))$.*

Then the sequence of products $A(k) \cdot A(k-1) \cdots A(1)$ converges to a stochastic matrix of rank 1.

The rest of this paper is devoted to the proof of Theorem 3.

2 Preliminaries

This section contains the basic tools for our convergence proof. On one hand, this is a semi-norm for stochastic matrices. We show that if the semi-norm of a product converges to zero, then firstly the product converges and secondly the limit matrix is stochastic of rank 1. On the other hand, we introduce a graph-theoretic interpretation of matrix products and present Wielandt's theorem, which we will use to show positivity of elements in partial products.

For denoting partial products, we introduce the notation

$$P(k) = A(k) \cdot A(k-1) \cdots A(1)$$

and

$$P(k, l) = A(k) \cdot A(k-1) \cdots A(l+1) .$$

In particular, $P(k) = P(k, l) \cdot P(l)$ whenever $l \leq k$.

2.1 Matrix Semi-Norm for Stochastic Matrices

All stochastic matrices have 1 as an eigenvalue of maximum modulus. If the matrix is irreducible, the corresponding right-eigenspace is one-dimensional and generated by the column vector $\mathbf{1} = {}^t(1, 1, \dots, 1)$. When studying such matrices, we are hence led to consider the distance of vectors to this eigenspace. Indeed, we will see in Lemma 2 that considering this distance is an appropriate tool for products of stochastic matrices. This technique for proving convergence results for products of stochastic matrices was communicated to the author by Charron-Bost [2].

Starting from the vector semi-norm that measures the distance to the common eigenspace $\mathbb{R} \cdot \mathbf{1}$, we will define a corresponding matrix semi-norm in analogy to the operator norm. A preliminary

result that we need later on is the following simple lemma which shows that the operator norm of a stochastic matrix with respect to the ∞ -norm is at most 1:

Lemma 1. *If A is an $n \times n$ stochastic matrix, then $\|Ax\|_\infty \leq \|x\|_\infty$ for all $x \in \mathbb{R}^n$.*

We define the vector semi-norm on \mathbb{R}^n by setting

$$\|x\| = \inf_{y \in \mathbb{R} \cdot \mathbf{1}} \|x - y\|_\infty .$$

This is indeed a semi-norm: The homogeneity property $\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$ follows directly from homogeneity of the ∞ -norm. The triangle inequality also holds: Let $x_1, x_2 \in \mathbb{R}^n$ and $y_1, y_2 \in \mathbb{R} \cdot \mathbf{1}$. The sum $y = y_1 + y_2$ remains in the eigenspace $\mathbb{R} \cdot \mathbf{1}$ and, because of the infimum in the definition of $\|x_1 + x_2\|$ and the triangle inequality for the ∞ -norm, we get

$$\|x_1 + x_2\| \leq \|x_1 + x_2 - y\|_\infty \leq \|x_1 - y_1\|_\infty + \|x_2 - y_2\|_\infty .$$

Forming the infimum over all y_1 and y_2 now shows $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$.

This vector semi-norm induces a matrix semi-norm on $\mathbb{R}^{n \times n}$ by defining it in the operator norm fashion:

$$\|A\| = \sup_{\substack{x \in \mathbb{R}^n \\ \|x\| \neq 0}} \frac{\|Ax\|}{\|x\|}$$

This definition ensures the homogeneity property $\|\alpha \cdot A\| = |\alpha| \cdot \|A\|$, the triangle inequality $\|A + B\| \leq \|A\| + \|B\|$, and submultiplicativity $\|A \cdot B\| \leq \|A\| \cdot \|B\|$. Clearly, $\|A\| = 0$ if the image of A is contained in the subspace $\mathbb{R} \cdot \mathbf{1}$.

The following lemma shows the utility of the matrix semi-norm to prove convergence of products of stochastic matrices. It is a core ingredient to our proof of Theorem 3.

Lemma 2. *If $\|P(k)\| \rightarrow 0$, then the sequence of partial products $P(k)$ converges and the limit is a stochastic matrix of rank 1.*

Proof. We show that, for every $x \in \mathbb{R}^n$, the sequence of vectors $P(k) \cdot x$ converges by showing that it is Cauchy. This then concludes the proof because stochasticity is preserved when taking limits and the matrix semi-norm is continuous, implying that the limit's semi-norm is zero.

Let $\varepsilon > 0$. Because also $\|P(k) \cdot x\| \rightarrow 0$, there exists some K such that $\|P(K) \cdot x\| \leq \varepsilon/2$. Letting $y \in \mathbb{R} \cdot \mathbf{1}$ such that $\|P(K) \cdot x\| = \|P(K) \cdot x - y\|_\infty$, we calculate for every $k \geq K$:

$$\begin{aligned} \|P(k) \cdot x - P(K) \cdot x\|_\infty &\leq \|P(k, K) \cdot P(K) \cdot x - y\|_\infty + \|P(K) \cdot x - y\|_\infty \\ &= \|P(k, K) \cdot (P(K) \cdot x - y)\|_\infty + \|P(K) \cdot x - y\|_\infty \\ &\leq 2 \cdot \|P(K) \cdot x - y\|_\infty = 2 \cdot \|P(K) \cdot x\| \leq \varepsilon \end{aligned}$$

because $P(k, K)$ is a stochastic matrix, which implies $y = P(k, K) \cdot y$ and that Lemma 1 is applicable. This shows that $P(k) \cdot x$ is a Cauchy sequence. \square

We now provide a tool to prove convergence of the matrix semi-norm of a product to zero by stating a sufficient condition for the semi-norm of a factor to be constantly bounded away from 1. It shows in particular that the semi-norm of a stochastic matrix is at most 1.

Lemma 3. *Let A be an $n \times n$ stochastic matrix with $A_{i,j} \geq \alpha$ for all $i, j \in [n]$. Then $\|A\| \leq 1 - n\alpha$.*

Proof. Write $A = E + B$ where $E = (\alpha)_{i,j}$. Because the image of E is contained in $\mathbb{R} \cdot \mathbf{1}$, we have $\|E\| = 0$ and hence $\|A\| = \|B\|$ by the triangle inequality. The entries of B are all nonnegative and its row sums $\sum_j B_{i,j}$ are all equal to $1 - n\alpha$.

Let $x \in \mathbb{R}^n$. For every $c \in \mathbb{R}$, we have:

$$\begin{aligned} \|Bx\| &\leq \|Bx - (1 - n\alpha)c \cdot \mathbf{1}\|_\infty = \max_i \left| \left(\sum_j B_{i,j} x_j \right) - (1 - n\alpha)c \right| \\ &= \max_i \left| \sum_j B_{i,j} \cdot (x_j - c) \right| \leq \max_i \sum_j B_{i,j} \cdot |x_j - c| \\ &\leq \max_i \sum_j B_{i,j} \cdot \|x - c \cdot \mathbf{1}\|_\infty = (1 - n\alpha) \cdot \|x - c \cdot \mathbf{1}\|_\infty \end{aligned}$$

Forming the infimum over all c shows $\|Bx\| \leq (1 - n\alpha) \cdot \|x\|$ and hence $\|A\| = \|B\| \leq 1 - n\alpha$. \square

2.2 Graph Interpretation of Matrix Products

Let i and j be nodes of a digraph G . A *walk* in G from i to j is a finite sequence of consecutive edges in G that starts at i and ends at j . Its length is the number of edges in the sequence. The empty sequence is a walk from i to j only if $i = j$.

The following lemma characterizes positivity of entries in products of stochastic matrices solely in terms of the matrices' associated digraphs. Denote by K_n the complete digraph with node set $[n]$.

Lemma 4. *Let $0 \leq l \leq k$ and $i, j \in [n]$. Then $P_{i,j}(k, l)$ is positive if and only if there exists a walk $(e_k, e_{k-1}, \dots, e_{l+1})$ of length $k - l$ in K_n from i to j such that e_m is an edge of $G(A(m))$ for all $l + 1 \leq m \leq k$.*

Proof. We proceed by induction on $k - l$.

The case $k = l$ is trivial because $P(k, k)$ is the identity matrix: $P_{i,j}(k, k)$ is positive if and only if $i = j$. But also, there exists a walk of length 0 from i to j if and only if $i = j$.

Let now $k - l \geq 1$. It is

$$P_{i,j}(k, l) = \sum_{i' \in [n]} A_{i,i'}(k) \cdot P_{i',j}(k-1, l) \quad (1)$$

by definition of $P(k, l)$.

First assume that $P_{i,j}(k, l)$ is positive. Then, by Equation (1), there exists some $i' \in [n]$ such that both $A_{i,i'}(k)$ and $P_{i',j}(k-1, l)$ are positive. By definition of $G(A(k))$, it contains the edge $e_k = (i, i')$. By the induction hypothesis, there exists a walk $(e_{k-1}, \dots, e_{l+1})$ in K_n from i' to j such that e_m is an edge of $G(A(m))$ for all $l + 1 \leq m \leq k - 1$. Hence the extended walk $(e_k, e_{k-1}, \dots, e_{l+1})$ fulfills the desired properties.

Now assume the existence of a walk $(e_k, e_{k-1}, \dots, e_{l+1})$ in K_n from i to j such that e_m is an edge of $G(A(m))$ for all $l + 1 \leq m \leq k$. Let $e_k = (i, i')$. By definition of $G(A(k))$, the entry $A_{i,i'}(k)$ is positive. Also, by the induction hypothesis, the entry $P_{i',j}(k-1, l)$ is positive. Hence, by Equation (1), also $P_{i,j}(k, l)$ is positive. \square

If a strongly connected digraph is aperiodic, there exist walks of arbitrary length between all pairs of nodes as long as the length is greater or equal to a number called the *exponent* (sometimes also *index*) of the digraph. Wielandt [5] provided an upper bound on the exponent. It is the best possible upper bound in terms of only the number of nodes.

Theorem 4 (Wielandt, 1950). *Let G be a strongly connected aperiodic digraph with n nodes and let i and j be nodes of G . Then there exists a walk in G from i to j of length k whenever k is greater or equal to*

$$W(n) = n^2 - 2n + 2 .$$

3 Proof of Theorem 3

Let $\alpha > 0$ be a lower bound on the positive matrix entries, i.e., $A_{i,j}(k) \geq \alpha$ whenever $A_{i,j}(k)$ is positive. In this section, we show the existence of a K such that

$$\|P(K)\| \leq 1 - n\alpha^{n(W(n)+1)} . \quad (2)$$

This suffices to show the theorem because the right-hand side of Equation (2) is strictly less than 1 and thus repeated application shows that $\|P(k)\| \rightarrow 0$ by submultiplicativity of the matrix semi-norm. Lemma 2 then concludes the proof.

For every $j \in [n]$, define $S_j(k)$ to be the set of indices $i \in [n]$ such that $P_{i,j}(k)$ is positive. We use Wielandt's theorem to show that the sequences $S_j(k)$ are almost nondecreasing:

Lemma 5. *Whenever $k - l \geq W(n)$, we have $S_j(l) \subseteq S_j(k)$.*

Proof. Let $i \in S_j(l)$. Denote by \hat{H} the strongly connected component of i in the aperiodic core H . Because \hat{H} is aperiodic, Theorem 4 implies the existence of a closed walk of length $k - l$ at node i in \hat{H} . The fact that \hat{H} is a subgraph of all $G(A(m))$ shows that $P_{i,i}(k, l)$ is positive by Lemma 4. Hence, because $P_{i,j}(l)$ is positive and $P_{i,j}(k) \geq P_{i,i}(k, l) \cdot P_{i,j}(l)$, so is $P_{i,j}(k)$. \square

Denote by $\mu_j(k)$ the smallest (positive) $P_{i,j}(k)$ where $i \in S_j(k)$. It is not hard to see that $\mu_j(k) \geq \alpha^{k-l} \cdot \mu_j(l)$ whenever $0 \leq l \leq k$. Under certain conditions we even have a stronger inequality:

Lemma 6. *If $S_j(k) = S_j(k+1)$, then $\mu_j(k+1) \geq \mu_j(k)$.*

Proof. Let $P_{i,j}(k+1)$ be positive, i.e., $i \in S_j(k+1) = S_j(k)$. By definition of $S_j(k)$, we have

$$P_{i,j}(k+1) = \sum_{\ell \in S_j(k)} A_{i,\ell}(k+1) \cdot P_{\ell,j}(k) . \quad (3)$$

Because $S_j(k) = S_j(k+1)$, we derive that $A_{i,\ell}(k+1)$ is zero whenever $i \notin S_j(k)$ and $\ell \in S_j(k)$. Hence, because $A(k)$ is completely irreducible, we also have that $A_{i,\ell}(k+1)$ is zero whenever $i \in S_j(k)$ and $\ell \notin S_j(k)$.

By assumption, we have $i \in S_j(k)$, and hence by the above and by stochasticity of $A(k+1)$:

$$1 = \sum_{\ell \in [n]} A_{i,\ell}(k+1) = \sum_{\ell \in S_j(k)} A_{i,\ell}(k+1) \quad (4)$$

Because $P_{\ell,j}(k) \geq \mu_j(k)$ for all $\ell \in S_j(k)$, combination of Equations (3) and (4) yields $P_{i,j}(k+1) \geq \mu_j(k)$. \square

Together with Lemma 3, the next lemma shows Equation (2), concluding the proof of Theorem 3.

Lemma 7. *There exists a K such that $P_{i,j}(k) \geq \alpha^{n(W(n)+1)}$ for all $i, j \in [n]$ and all $k \geq K$.*

Proof. Let $j \in [n]$. For every $i \in [n]$, let k_i be the least nonnegative integer such that $i \in S_j(k_i)$. Because of the eventual positivity property (2) in the theorem statement, all k_i are well-defined. By permutating indices, we can assume without loss of generality that $k_1 \leq k_2 \leq \dots \leq k_n$. Because $P(0)$ is the identity matrix, we have $S_j(0) = \{j\}$ and hence $k_1 = 0$.

We inductively show

$$\mu_j(k_m) \geq \alpha^{(m-1)(W(n)+1)} \quad (5)$$

for all $1 \leq m \leq n$. This is true for $m = 1$. To prove the inductive step, we distinguish two cases:

(A) $k_m - k_{m-1} < W(n)$ and (B) $k_m - k_{m-1} \geq W(n)$.

In case (A), we have

$$\mu_j(k_m) \geq \alpha^{k_m - k_{m-1}} \cdot \mu_j(k_{m-1}) \geq \alpha^{(m-1)(W(n)+1)}$$

by the induction hypothesis.

In case (B), we have $S_j(k) = S_j(k_{m-1})$ for all k with $k_{m-1} + W(n) \leq k \leq k_m - 1$ by Lemma 5 and the definition of k_m . Repeated application of Lemma 6 hence yields $\mu_j(k_m - 1) \geq \mu_j(k_{m-1} + W(n))$. We thus have

$$\begin{aligned} \mu_j(k_m) &\geq \alpha \cdot \mu_j(k_m - 1) \geq \alpha \cdot \mu_j(k_{m-1} + W(n)) \\ &\geq \alpha^{W(n)+1} \cdot \mu_j(k_{m-1}) \geq \alpha^{(m-1)(W(n)+1)} \end{aligned}$$

by the induction hypothesis.

In particular, we have shown Equation (5) for $m = n$. Now set $K = k_n + W(n)$. By Lemma 5, $S_j(k) = [n]$ for all $k \geq K$. Also, $\mu_j(k) \geq \mu_j(K) \geq \alpha^{W(n)} \cdot \mu_j(k) \geq \alpha^{n(W(n)+1)}$ by Lemma 6 and Equation (5). This shows the entries of the j th column of $P(k)$ are all greater or equal to $\alpha^{n(W(n)+1)}$ whenever $k \geq K$. Because the choice of j was arbitrary, we have shown the lemma. \square

References

- [1] Vincent D. Blondel, Julien M. Hendrickx, Alex Olshevsky, and John N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference (CDC-ECC)*, pp. 2996–3000. IEEE, New York, NY, 2005.
- [2] Bernadette Charron-Bost. Personal communication, 2012.
- [3] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE* 95(1):215–233, 2007.
- [4] Behrouz Touri and Angelia Nedić. Product of random stochastic matrices. [arXiv:1110.1751v1](https://arxiv.org/abs/1110.1751v1) [math.PR], 2011.
- [5] Helmut Wielandt. Unzerlegbare, nicht negative Matrizen. *Mathematische Zeitschrift* 52(1):642–648, 1950.
- [6] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the AMS* 14(5):733–737, 1963.